DOCUMENT RESUME

ED 316 581 TM 014 553

AUTHOR Young, John W.
TITLE A General Linear Model Approach to Adjusting the Cumulative GPA.
PUB DATE 90
NOTE 17p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Ability Identification; *Admission Criteria; *College Admission; *Grade Point Average; Higher Education; Least Squares Statistics; Mathematical Models; Matrices; *Multiple Regression Analysis; Predictive Validity; *Predictor Variables
IDENTIFIERS *General Linear Model; Scholastic Aptitude Test; Stanford University CA

ABSTRACT

A general linear model (GLM), using least-squares techniques, was used to develop a criterion measure to replace freshman year grade point average (GPA) in college admission predictive validity studies. Problems with the use of GPA include those associated with the combination of grades from different courses and disciplines into a single measure, the incompatibility of grades across courses, and adjustments for differences in grading standards at the course level. Freshman year GPA has the added disadvantage of being an insufficient and inadequate sample of a student's achievement through four years of college. Data for this study were taken from all 1,564 students who entered Stanford University (California) as freshmen in the autumn of 1982. Data include: sex, ethnicity, scores on the verbal and mathematics tests of the Scholastic Aptitude Test, high school GPA, high school rank in class, and data on courses taken at Stanford University. Estimates of each student's ability were obtained based on performance in college courses representing each of three substantive areas. Ability estimation was accomplished using the PROC GLM procedure in Statistical Analysis System. The data matrix for each area consisted of students as rows and courses as columns, with each cell of the matrix representing a student's grade in a particular course. The GLM produces an estimate of the least-squares mean for each student, which is the mean grade that would have resulted if the data matrix was complete. Ability estimates are used as separate outcome measures in a series of predictive validity studies and compared with the usually computed GPA. Multiple regression analysis was conducted using other traditional preadmissions measures. Results indicate that this approach yields a predictable criterion. Fifteen data tables are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED316581

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John W. Young
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A General Linear Model approach to adjusting the cumulative GPA

John W. Young, Ph.D.
Dept. of Educational Psychology
Rutgers University

Introduction

Criterion-related validity is a central concept in educational and psychological measurement. We are often interested in the relationship of a given predictor or set of predictors to some outcome measure. That is, how valid or how useful are these predictors in estimating some criterion? In institutional research on undergraduate admissions, the predictive validity of preadmissions measures such as standardized test scores and high school grades in determining academic performance in college is of interest to students, admissions officers, university administrators, educational researchers, guidance counselors, and the general public.

One limitation of many predictive validity studies that examined the utility of preadmissions measures is the use of the freshman year grade point average (GPA) as the criterion (Wilson, 1983). The freshman year GPA is favored in institutional studies because it is a well-defined criterion, it is easily obtainable from university records, and it is available relatively soon after the matriculation of a class of students. However, the

The SAS code written for this study can be obtained by writing to: Prof. John W. Young, Dept. of Educational Psychology, Graduate School of Education, Rutgers University, New Brunswick, NJ 08903.

TM014553

BEST COPY AVAILABLE

freshman year GPA is neither a sufficient nor adequate sample of a student's achievement through four years of college.

Within any university, grading standards can vary from instructor to instructor and from department to department, creating a problem known as 'differential departmental grading' (see e.g., Goldman & Hewitt, 1975; Willingham, 1985). The net effect of this problem is to diminish the construct validity of the GPA scale. In addition, the GPA is a multidimensional measure of performance condensed down to a single number. By distilling grades from a variety of departments into a single composite, we insure that the GPA will be an unreliable criterion.

Previously, most of the criticisms about the effectiveness of preadmissions measures has focused on the limitations of the predictors. However, the controversy about validity coefficients has not properly taken into account the fact that the GPA criterion (both first year and cumulative) has certain correctable defects. One of the basic facts in measurement is that a variable with significant measurement error will have artificially depressed correlation coefficients of that variable with any other measure (Cronbach, 1984).

Other criteria used in predictive validity studies have included grades in individual courses (McCornack & McLeod, 1988). Individual course grades, however, are difficult to predict and are theoretically less reliable than a composite based on several grades. Alternatively, Elliot & Strenta (1988) examined differential validity by adjusting GPA's (for individual years and cumulatively) using a simple linear correction by departments.

This study extends the current research into predictive validity in three ways. First, most GPA's as currently computed are multidimensional

measures of academic performance since grades from different courses and disciplines are combined into a single measure. It is unreasonable to assume that grades from all of the courses at a university are determined by a single ability construct. One solution would be to partition courses into relatively unidimensional subsets and compute a separate GPA for each set of courses. Each of these GPA's would then be a better defined and more reliable composite than the usually computed GPA.

Second, grades across courses are not directly comparable. Differences in grading standards among departments and instructors is the primary source of unreliability when using an unadjusted cumulative GPA as an outcome measure (see e.g., Goldman & Hewitt, 1975; Goldman & Slaughter, 1976; Willingham, 1985). Previous attempts to adjust grades have been marginally successful (see e.g., Linn, 1966; Boldt, 1976; Elliot & Strenta, 1988). Item Response Theory (IRT) has also been used to equate grades across courses (Young, 1988, 1989a, 1989b, 1989c, 1990). An approach using the General Linear Model (GLM) to equate grades from different courses within an institution appears to be theoretically promising but has not been empirically tested.

Third, adjustments for differences in grading standards should be made at the course level rather than at the department level in order to adequately detect differential predictive validity. Adjusting GPA's for between-department differences in grading alone masks differences among courses within the same department (Elliot & Strenta, 1988). Information about grades at the course level is needed to determine individual student performance.

This study attempts to address all three of these problems by

developing a new criterion measure using the General Linear Model.

The General Linear Model

GLM using classical least-squares techniques is the most powerful of all statistical models (see e.g., Searle, 1971). In the context of this study, GLM takes the form of an incomplete design, specifically a chain block design (Youden & Connor, 1953). Multiple measurements for each student (i.e., block) is available in the form of course grades. Each unique college course is considered the equivalent of a treatment. An estimate of the effect due to each course can be obtained and used to compute an adjusted cumulative GPA for each student.

For this study, the GLM is an additive model with main effects only and has the following form:

$$GRADE_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} ,$$

where $GRADE_{ij}$ is the numerical value of the letter grade for the i-th student in the j-th course, μ is the grand mean of course grades, α_i is the 'effect' of the i-th student, β_j is the 'effect' of the j-th course, and ϵ_{ij} is the error term. Quotations are used around the word effect since this study does not satisfy the requirements for a true experimental study and no causal mechanism is postulated. Note that in this model, there is no interaction term, $(\alpha\beta)_{ij}$, since each student only takes each course once. Thus, estimation of this parameter is not possible since, in this model, it is absorbed by the error term.

The sample estimates of the parameters for the GLM are given by:

$$\widehat{GRADE}_{ij} = \bar{X}.. + \hat{\alpha}_i + \hat{\beta}_j .$$

To compute an adjusted GPA for each student, we need an estimate of the

average grading 'effect' due to each student. This estimate is given by the following expression:

$$\hat{\alpha}_i = \sum_j (\widehat{GRADE}_{ij} - \bar{X}.. - \hat{\beta}_j) / j .$$

It should be noted that the students by courses data matrix in this study does not meet some of the assumptions for a randomized design. Random assignment of students to courses is impossible for obvious reasons. In addition, the assumption of no student-by-course interactions may be violated since a student's grade may be determined by factors other than course performance, such as judicious course selection. Nevertheless, the advantage of this approach is that it enables us to develop a statistical procedure for adjusting a student's GPA.

Methodology

Data for this study originated at Stanford University. The sample is defined as all students who entered the university as freshmen in Autumn 1982, nominally the Class of 1986. The sample size is 1564 and includes students who left before obtaining a degree. For each student, the following information is available: sex, ethnicity, SAT-V and SAT-M scores, HS GPA, HS rank in class, and data on courses taken at Stanford. Using a minimum course enrollment of 15 students from this cohort, courses were divided into three substantive areas: 396 natural science and engineering courses, 184 humanities courses, and 198 social science courses, for a total of 778 courses across the four academic years.

Estimates of each student's ability was obtained based on performance in courses representing each of the substantive areas. Ability estimation was accomplished using the PROC GLM procedure in SAS (SAS Institute Inc.,

1985). The data matrix consisted of students as rows and courses as columns with each cell of the matrix representing a student's grade in a particular course, and is relatively empty since each student has only taken a small fraction of the available courses. In fact, the three data matrices (one for each of the substantive areas) have non-missing values in about 3-4% of the cells. Only letter grades were used in this study and were converted to numerical values using the Stanford registrar's grade scale (i.e., A=4, B=3, C=2, D=1; plus/minus suffixes modify the letter grade $\pm$0.3).

The statistical estimation problem here can be characterized as one of an incomplete design with a large degree of missing data. PROC GLM produces an estimate of the least-squares mean for each student which is the mean grade that would have resulted if the data matrix was complete. The least-squares means, then, are the GPA's hypothetically based on every student having taken every course, and will be referred to in this paper as $LS\text{-}GPA_{ik}$ ($i=1,\ldots,1564$, $k=1,2,3$). These ability estimates were then used as separate outcome measures in a series of predictive validity studies and compared with the usually computed GPA. Multiple regression analysis was conducted using the three traditional preadmissions measures of SAT-V, SAT-M, and HS GPA as predictors. In addition, the three LS-GPA's will be combined into a single composite using canonical correlation analysis and used in further analyses as a single criterion ($LS\text{-}GPA_{i.}$).

Results

Preliminary results indicate that this approach appears to yield a more predictable criterion than the usual GPA. Results comparing the LS-GPA and the GPA are given on pages 9-11 (Figures 1-3). Note, however, that although

the statistically adjusted LS-GPA may be more predictable than the GPA, this does not necessarily imply that it is a more valid measure of college performance than the cumulative GPA. Evidence on the construct validity of each of the two measures is needed in order to support or refute that particular hypothesis.

Ideally, one would like to obtain a LS-GPA for each student who has taken at least two courses in each of the three substantive areas. This requires substantial computing resources because building the students by courses data matrix is a highly computing intensive task. However, because of the present hardware configuration at Rutgers University, there is a limitation on the size of the matrix that can be analyzed. The computer used was a National Advanced Systems (NAS) AS/9000-2 mainframe (comparable to an IBM 3033 mainframe). This problem can be overcome by utilizing a more powerful computer with greater virtual memory or by rewriting the SAS program in a manner that uses virtual memory more efficiently.

Preliminary results are based on samples of students who have taken more courses than the ideal minimum of two. Analysis of the natural science and engineering courses was restricted to students who have taken at least 24 courses out of the total pool of 396 courses. This yielded a sample size of 223. Since Stanford University operates on a calendar of three quarters per academic year, this averages out to two courses in this area per quarter over four years for students in the sample. Analysis of the social science courses was restricted to students who took a minimum of 8 courses (N=372). The same minimum of 8 courses was also used for the humanities area (N=388).

There is some overlap between the two samples of students who took the minimum number of courses in the humanities and in the social science areas.

There are few students who are in the sample for natural science and engineering courses and also one of the other samples. The sample sizes given in Figures 1-3 are slightly smaller than the ones listed here because a few students were missing either SAT scores or HS GPA.

In the GLM with students and courses as main 'effects', we find that the proportion of variance explained (R^2) in natural science and engineering grades by the two sources is .57. In comparison, the R^2 for social science grades is .62; for humanities grades, the R^2 is .56. For these analyses, the students by courses data matrices have non-missing values in about 5-6% of the cells. Obtaining LS-GPA estimates for samples of about 400 students required around 12 minutes of CPU on the NAS AS/9000-2.

FIGURE 1

Sample: Students with grades in 24+ natural science and eng. courses (N=216)

Dependent Variable: GPA based on 396 natural science and eng. courses

Analysis of Variance Source Table

Source	df	SS	MS	F-ratio	R^2
Model	3	13.87	4.62	25.31	.2637
Error	212	38.72	.1826		
Total	215	52.59			

Multiple Regression Analysis Parameter Estimates

Variable	Estimate	Std Error	t-statistic	p <
intercept	-1.3311	0.6601	-2.02	.0450
SAT-V	1.28E-3	4.69E-4	2.73	.0068
SAT-M	2.52E-3	5.42E-4	4.64	.0001
HS GPA	0.5060	0.1673	3.03	.0028

Dependent Variable: LS-GPA based on 396 natural science and eng. courses

Analysis of Variance Source Table

Source	df	SS	MS	F-ratio	R^2
Model	3	17.18	5.73	29.59	.2952
Error	212	41.03	.1935		
Total	215	58.21			

Multiple Regression Analysis Parameter Estimates

Variable	Estimate	Std Error	t-statistic	p <
intercept	-1.5949	0.6795	-2.35	.0198
SAT-V	1.41E-3	4.83E-4	2.92	.0039
SAT-M	2.94E-3	5.58E-4	5.26	.0001
HS GPA	0.4987	0.1722	2.90	.0042

Correlation Coefficient Matrix

	LS-GPA	GPA	SAT-V	SAT-M	HS GPA
LS-GPA	1.00				
GPA	.99	1.00			
SAT-V	.40	.38	1.00		
SAT-M	.48	.44	.48	1.00	
HS GPA	.29	.29	.21	.21	1.00

FIGURE 2

Sample: Students with grades in 8+ social science courses (N=362)

Dependent Variable: GPA based on 198 social science courses

Analysis of Variance Source Table

Source	df	SS	MS	F-ratio	R^2
Model	3	37.69	12.56	77.87	.3949
Error	358	57.77	.1614		
Total	361	95.46			

Multiple Regression Analysis Parameter Estimates

Variable	Estimate	Std Error	t-statistic	p <
intercept	-0.9340	0.3040	-3.07	.0023
SAT-V	1.67E-3	3.20E-4	5.22	.0001
SAT-M	2.42E-3	3.44E-4	7.04	.0001
HS GPA	0.4052	0.0768	5.28	.0001

Dependent Variable: LS-GPA based on 198 social science courses

Analysis of Variance Source Table

Source	df	SS	MS	F-ratio	R^2
Model	3	44.07	14.69	87.32	.4225
Error	358	60.23	.1683		
Total	361	104.31			

Multiple Regression Analysis Parameter Estimates

Variable	Estimate	Std Error	t-statistic	p <
intercept	-1.2083	0.3105	-3.89	.0001
SAT-V	1.84E-3	3.26E-4	5.63	.0001
SAT-M	2.65E-3	3.51E-4	7.56	.0001
HS GPA	0.4141	0.0784	5.28	.0001

Correlation Coefficient Matrix

	LS-GPA	GPA	SAT-V	SAT-M	HS GPA
LS-GPA	1.00				
GPA	.98	1.00			
SAT-V	.52	.49	1.00		
SAT-M	.55	.53	.52	1.00	
HS GPA	.38	.38	.27	.25	1.00

FIGURE 3

Sample: Students with grades in 8+ humanities courses (N=377)

Dependent Variable: GPA based on 184 humanities courses

Analysis of Variance Source Table

Source	df	SS	MS	F-ratio	R^2
Model	3	11.98	3.99	45.30	.2670
Error	373	32.89	.0881		
Total	376	44.88			

Multiple Regression Analysis Parameter Estimates

Variable	Estimate	Std Error	t-statistic	p <
intercept	0.9861	0.2163	4.56	.0001
SAT-V	1.44E-3	2.30E-4	6.26	.0001
SAT-M	6.89E-4	2.36E-4	2.92	.0037
HS GPA	0.2821	0.0518	5.45	.0001

Dependent Variable: LS-GPA based on 184 humanities courses

Analysis of Variance Source Table

Source	df	SS	MS	F-ratio	R^2
Model	3	11.55	3.85	47.47	.2763
Error	373	30.25	.0811		
Total	376	41.79			

Multiple Regression Analysis Parameter Estimates

Variable	Estimate	Std Error	t-statistic	p <
intercept	1.1832	0.2074	5.71	.0001
SAT-V	1.75E-3	2.21E-4	7.91	.0001
SAT-M	2.18E-4	2.26E-4	0.97	.3347
HS GPA	0.2672	0.0497	5.38	.0001

Correlation Coefficient Matrix

	LS-GPA	GPA	SAT-V	SAT-M	HS GPA
LS-GPA	1.00				
GPA	.92	1.00			
SAT-V	.46	.42	1.00		
SAT-M	.26	.33	.41	1.00	
HS GPA	.35	.35	.23	.23	1.00

Discussion

The results indicate that the increase in predictability from using the LS-GPA is largest in the natural science and engineering area, and smallest in the humanities area. This is consistent with the findings obtained from using Item Response Theory to adjust college grades (Young, 1989a, 1989b, 1989c). Changes in predictability is measured by comparing the difference in proportion of variance explained (R^2) in each of the criterion using the three predictors of SAT-V, SAT-M, and HS GPA in multiple regression analyses.

Certain changes in the relative contribution of each of the predictors in the regression equations should be noted. SAT-M scores are somewhat better in predicting the natural science and engineering LS-GPA than of the unadjusted GPA (see Figure 1). Since many natural science and engineering courses require a high degree of quantitative reasoning ability, we might expect that SAT-M scores should be a very good predictor of grades. This is more evident when the criterion used is the LS-GPA. In comparison, for the humanities area, SAT-V scores predicted the LS-GPA better than the GPA. Additionally, SAT-M is not statistically significant in the regression equation for predicting humanities grades when we use the LS-GPA as the criterion instead of the usual GPA (see Figure 3). This result tends to support the hypothesis that verbal aptitude is much more important than quantitative aptitude for predicting performance in humanities courses. No improvement in the predictive ability of the HS GPA was found in this study. We expect that SAT scores would show improvement as predictors, but not HS GPA, because SAT scores are highly reliable measures. HS GPA, on the other hand, is a multidimensional composite and suffers from the same drawbacks as

the usual college GPA.

There are several statistical reasons that may help to explain the finding that the largest observed improvement occurs in the natural science and engineering area. First, the distributions of grades in the natural science and engineering courses are more variable. The problem of grade inflation where all or almost all of the students in a course receive A's is especially true for many of the humanities courses. Second, the precision of the LS-GPA estimates is best for the natural science and engineering courses. This is due to the fact that more course grades are used to compute the natural science and engineering LS-GPA.

Estimation of performance based on the humanities courses is the most difficult of the three cognate areas. The underlying ability trait in humanities courses may be less well defined or identifiable than in the other areas. One possible explanation for this is that there may be a wider variety of skills needed to perform well in humanities courses. Courses in this area may differ from each other to a greater extent than in the natural science and engineering or social science areas. Prediction of performance for social science courses falls between the natural science and engineering and humanities courses. This seems to be consistent with the substantive and statistical comparisons among the three subsets of courses.

It is expected that additional analyses based on samples which include students who have taken even fewer courses than in the results given here will indicate an even greater improvement in predictability of the LS-GPA over the usual GPA. One reason for this expectation is that greater observed variability in the criterion should occur with larger samples.

Summary

The main contribution of this study is to demonstrate that research on this topic can benefit from developing measures of academic performance that are more unidimensional and that can also account for differences in grading standards among courses. It appears that the use of this statistical methodology can create a more predictable criterion, the LS-GPA, that is less contaminated with some of the measurement error that may be inherent in the grading process. This new outcome measure will hopefully lead to a better understanding of the true predictive power of preadmissions information in forecasting academic performance.

References

Boldt, R. F. (1976) Efficacy of undergraduate grade adjustment for improving the prediction of law school grades (Report No. LSAC 76-4). In Law School Admissions Council, Reports of LSAC Sponsored Research: Vol. III, 1975-1977. Princeton, NJ: Law School Admissions Council.

Cronbach, L. J. (1984) Essentials of psychological testing (4th ed.). New York: Harper & Row.

Elliot, R. & Strenta, A. C. (1988) Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. Journal of Educational Measurement, 25, 333-347.

Goldman, R. D. & Hewitt, B. N. (1975) Adaptation-level as an explanation for differential standards in college grading. Journal of Educational Measurement, 12, 149-161.

Goldman, R. D. & Slaughter, R. E. (1976) Why college grade point average is difficult to predict. Journal of Educational Psychology, 68, 9-14.

Linn, R. L. (1966) Grade adjustments for prediction of academic performance: A review. Journal of Educational Measurement, 3, 313-329.

McCornack, R. L. & McLeod, M. M. (1988) Gender bias in the prediction of college course performance. Journal of Educational Measurement, 25, 321-331.

SAS Institute Inc. (1985) SAS user's guide: Statistics, version 5 edition. Cary, NC: SAS Institute Inc.

Searle, S. R. (1971) Linear models. New York: John Wiley.

Willingham, W. W. (1985) Success in college. New York: College Entrance Examination Board.

Wilson, K. M. (1983) A review of research on the prediction of academic performance after the freshman year. College Board Report No. 83-2 and Educational Testing Service Research Report No. 83-11. New York: College Entrance Examination Board.

Youden, W. J. & Connor, W. S. (1953) The chain block design. Biometrics, 9, 127-140.

Young, J. W. (1988) Development of a universal grade scale using Masters' Partial Credit Model. Paper presented at the annual meeting of the Psychometric Society, Los Angeles, CA, June 1988.

Young, J. W. (1989a) Developing a universal scale for grades: Investigating predictive validity in college admissions (Doctoral dissertation, Stanford University, 1988).

Young, J. W. (1989b) Adjusting the cumulative GPA using Item Response Theory. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, March 1989.

Young, J. W. (1989c) Adjusting the cumulative GPA using Item Response Theory. Journal of Educational Measurement, in press.

Young, J. W. (1990) Are validity coefficients understated due to correctable defects in the GPA? Journal of Higher Education, under review.